

SENTIMENT ANALYSIS IN TWITTER

A Project Report

*Submitted in the partial fulfillment of the requirement
for the award of the degree of*

**BACHELOR OF ENGINEERING
IN
INFORMATION TECHNOLOGY
BY**

RAJARSHI SARKAR (BE/1397/2011)

AMIT KUMAR (BE/1513/2011)

**UNDER GUIDANCE OF
DR. VIJAY KUMAR JHA
ASSOCIATE PROFESSOR**



**DEPT. OF COMPUTER SCIENCE AND ENGINEERING
BIRLA INSTITUTE OF TECHNOLOGY
MESRA, RANCHI – 835215**

DECLARATION CERTIFICATE

This is to certify that the work presented in the project report entitled “**Sentiment Analysis in Twitter**” in partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering in Information Technology of Birla Institute of Technology, Mesra, Ranchi is an authentic work carried out under my supervision and guidance.

To the best of my knowledge, the content of this project report does not form a basis for the award of any previous degree to anyone else.

Date:

Dr. Vijay Kumar Jha
Dept. of Computer Science & Engineering
Birla Institute of Technology
Mesra, Ranchi-835215

CERTIFICATE OF APPROVAL

The foregoing project entitled “**Sentiment Analysis in Twitter**”, is hereby approved as a creditable study of research topic and has been presented in satisfactory manner to warrant its acceptance as prerequisite to the degree for which it has been submitted.

It is understood that by this approval, the undersigned do not necessarily endorse any conclusion drawn or opinion expressed therein, but approve the project report for the purpose for which it is submitted.

(Internal Examiner)

(External Examiner)

(Dr. Sandip Dutta)

Head of the Department

Dept. of Computer Science & Engineering,

Birla Institute of Technology,

Mesra, Ranchi-835215

ACKNOWLEDGEMENT

We owe our deepest gratitude to our guide **Dr. Vijay Kumar Jha**, who helped us throughout the project. He made sure that we learnt by practice, always motivating us to think a step further, work a little harder. Our interactions with him always resulted in newer ideas and proved beneficial towards our work. Without his constant presence and supervision our work would not have been successful.

We especially acknowledge the many useful discussions we had among ourselves that helped us understand some of the subtle technical problems in a better way.

Rajarshi Sarkar

(BE/1397/2011)

Amit Kumar

(BE/1513/2011)

ABSTRACT

Twitter is a popular microblogging service where users create status messages (called "tweets"). These tweets sometimes express opinions about different topics. Sentiment Analysis in Twitter is a method to automatically extract sentiment (positive or neutral or negative) from a tweet. This is very useful because it allows feedback to be aggregated without manual intervention. Consumers can use sentiment analysis to research products or services before making a purchase. Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. Organizations can also use this to gather critical feedback about problems in newly released products. There has been a large amount of research in the area of sentiment classification. Traditionally most of it has focused on classifying larger pieces of text, like reviews. Tweets (and microblogs in general) are different from reviews primarily because of their purpose: while reviews represent summarized thoughts of authors in a specific topic, tweets are short messages about a variety of topics and are limited to 140 characters of text. Also, the frequency of misspellings, acronyms and slang in tweets is much higher than in other domains.

TABLE OF CONTENTS

1. Introduction	1-2
1.1. Defining the sentiment	2
1.2. Characteristic of Tweets	2
2. Why Sentiment Analysis in Twitter?	3
3. Objective and Scope of the Project	3
4. Tools used	4
5. Fetching Tweets	5
6. Preprocessing Tweets	6-7
7. Approaches	8-19
7.1. Sentiment Dictionary Approach	9-11
7.1.1. Overview	9
7.1.2. Algorithm	9-10
7.1.3. Difficulties faced	10
7.1.4. Results	11
7.2. Naïve Bayes Approach	12-14
7.2.1. Overview	12
7.2.2. Algorithm	13
7.2.3. Difficulties faced	14
7.2.4. Results	14
7.3. Support Vector Machines Approach	15-19
7.3.1. Overview	15-17
7.3.2. Algorithm	18
7.3.3. Difficulties faced	18
7.3.4. Results	19
8. Future work	20
9. Conclusion	21
10. Appendix	22-61
10.1. Appendix A: Sentiment Dictionary Approach Codes	22-27
10.2. Appendix B: Naïve Bayes Approach Codes	28-53
10.3. Appendix C: Support Vector Machines Codes	54-61
11. References	62-63